Density Estimation in the Uniform Norm and White Noise Approximation

Alexander Korostelev[1] and Michael Nussbaum
Wayne State University, Detroit and
Weierstrass Institute, Berlin

June 1995

Abstract

We develop the exact constant of the risk asymptotics in the uniform norm for density estimation. This constant has first been found for nonparametric regression and for signal estimation in Gaussian white noise. We show that for densities with Hölder exponent $> 1/2$, the formal approximation of the i. i. d. experiment by Gaussian white noise in the sense of Le Cam's deficiency distance (asymptotic equivalence) can be utilized. For densities with Hölder exponent $\leq 1/2$ where asymptotic equivalence fails, the result can still be established independently.

1 Introduction

Recently in [5] an asymptotically minimax exact constant has been found for loss in the uniform norm, for Gaussian nonparametric regression when the parameter set is a Hölder function class. Donoho [3] subsequently extended the result to signal estimation in Gaussian white noise and showed it to be related to nonstochastic optimal recovery. This risk bound represents an analog of the now classical L_2-minimax constant of Pinsker [10] valid for a Sobolev function class. From a risk bound valid in white noise, abstract decision theory allows to deduce asymptotic risk bounds for other models, by reference to the concept of asymptotic equivalence of experiments (in the sense of

---

[1] Research supported in part by the International Science Foundation under Grant MCF000 and by the Deutsche Forschungsgemeinschaft under Grant SFB 373.
1990 *Mathematics Subject Classification.* Primary 62 G 07 ; Secondary 62 B 15, 62 G 20
Key words and phrases. Density estimation, uniform norm loss, asymptotic minimax risk, exact constant, nonparametric experiments, Gaussian white noise, deficiency distance, asymptotic equivalence.

Le Cam's deficiency distance). For density estimation this asymptotic equivalence is developed in [9]; examples of application related to the Pinsker bound can be found there. But this general reduction to Gaussian white noise, for all decision problems with uniformly bounded loss, is not possible if the parameter space is too large; indeed the smoothness index 1/2 has been shown to be critical (Brown and Zhang [1]). The purpose of this note is to discuss the potential and the limits of the asymptotic equivalence approach, in the context of the sup-norm minimax constant for density estimation.

Consider a sample $X_1, ..., X_n$ of i. i. d. observations having a probability density $f = f(x)$ in the interval $0 \le x \le 1$. Let β, L and b be some positive constants, $0 < b < 1$, and let $\Sigma(\beta, L, b)$ be the class of Hölder densities

$$\begin{aligned}\Sigma(\beta, L, b) \;=& \Big\{ g : \int_0^1 g = 1,\; g(x) \ge b \text{ for } 0 \le x \le 1, \text{ and} \\ & |g^{\lfloor\beta\rfloor}(x_1) - g^{\lfloor\beta\rfloor}(x_2)| \le L|x_1 - x_2|^{\beta-\lfloor\beta\rfloor}, \quad 0 \le x_1, x_2 \le 1 \Big\}.\end{aligned}$$

where $\lfloor\beta\rfloor$ denotes the largest integer strictly less than β. Assume that the density f belongs a priori to $\Sigma(\beta, L, b)$. Consider an arbitrary estimator $\hat{f}_n = \hat{f}_n(x)$ measurable w.r.t. the observations $X_1, ..., X_n$. We define the discrepancy of $\hat{f}_n(x)$ and the true density $f(x)$ by the sup–norm $||\hat{f}_n - f||_\infty$ where

$$||f||_\infty \;=\; \sup_{0\le x\le 1} |f(x)|.$$

Denote by $P_f^{(n)}$ the probability distribution of the observations $X_1, ..., X_n$, and by $E_f^{(n)}$ the expectation w.r.t. $P_f^{(n)}$. Let $w(u),\ u \ge 0$, be a continuous bounded loss function, i.e. a monotone function of u, $w(0) = 0$. Introduce the minimax risk

(1) $$r_n = r_n(w(\cdot); \beta, L, b) = \inf_{\hat{f}_n} \sup_{f \in \Sigma(\beta,L,b)} E_f^{(n)} w(\psi_n^{-1} ||\hat{f}_n - f||_\infty)$$

where $\psi_n = ((\log n)/n)^{\beta/(2\beta+1)}$. This normalization factor guarantees a non–degenerate behaviour of the risk (1) as $n \to \infty$, cp. Ibragimov and Khasminskii [4], Devroye and Györfi [2]. The goal of this paper is to find the exact asymptotics of the risk (1). To do this we need two additional definitions. First, note that the densities in $\sum(\beta, L, b)$ are unifomly bounded, i.e.

(2) $$B_* = B_*(\beta, L, b) = \max_{f\in\Sigma(\beta,L,b)} \max_{0\le x\le 1} f(x) < +\infty\,.$$

Secondly, denote by $\Sigma_0(\beta)$ an auxiliary class of Hölder functions on the whole real line with constant $L = 1$:

$$\Sigma_0(\beta) = \Big\{ g(x),\, x \in R^1 \;: |g^{\lfloor\beta\rfloor}(x_1) - g^{\lfloor\beta\rfloor}(x_2)| \le |x_1 - x_2|^{\beta-\lfloor\beta\rfloor},\, x_1, x_2 \in R^1 \Big\}.$$

Let $||g||_2$ denote the L_2-norm of g. Define the constant

(3) $$A_\beta = \max \Big\{ g(0) \mid \quad ||g||_2 \le 1,\; g \in \Sigma_0(\beta) \Big\}.$$

Theorem 1. *For any $\beta > 1/2$ and for any loss function $w(u)$ the following equality holds for the minimax risk (1):*

$$\lim_{n\to\infty} r_n = w\left(A_\beta \left(\frac{2B_* L^{1/\beta}}{2\beta+1} \right)^{\beta/(2\beta+1)} \right) \tag{4}$$

where $B_* = B_*(\beta, L, b)$ and A_β are defined by (2) and (3) respectively.

In the next section we will see how this result can be deduced from the one in a Gaussian continuous regression model, via asymptotic equivalence. We will then show that although equivalence fails for $\beta < 1/2$, the present minimax risk asymptotics still holds for any smoothness index $\beta > 0$ in the density model, as it does in continuous regression. For estimation problems which are local on $[0, 1]$, like estimation of a density at a point, such a result would not be surprising and can in fact be deduced from white noise using an appropriate concept of local asymptotic equivalence; see Low [8]. The minimax risk for the uniform norm, however, although it is related to kernel smoothing of data on $[0, 1]$, is a nonlocal one with respect to the interval. Indeed it can be shown that when the sup-norm loss is taken over a shrinking subinterval which is shrinking slower than the pertaining bandwidth, then the same minimax rate applies but the minimax constant is smaller. Nevertheless the global sup-norm minimax constant is valid in both the density and the white noise model for low smoothness indices β near 0 where the global asymptotic equivalence of the two experiments fails. This underlines the need for further study of concepts of reduced equivalence, pertaining to restricted classes of decision problems only, as suggested by L. Le Cam in [7] and [6].

2 Equivalence of Density and Regression Experiments

We obtain the proof of Theorem 1 reducing the problem to that in nonparametric regression via the equivalence of statistical experiments, see [9]. For any $f_0 \in \Sigma(\beta, L, b)$ and for $n \geq 3$ introduce the neighborhood of f_0 by

$$U_n = U_n(f_0) = \left\{ f : \, f \in \Sigma(\beta, L, b), \, ||f - f_0||_\infty \leq (n^{1/4} \log n)^{-1} \right\}. \tag{5}$$

For any fixed $f_0 \in \sum(\beta, L, b)$ consider a regression problem with observation $Y^{(n)} = Y^{(n)}(x)$ satisfying the Ito equation

$$\dot{Y}^{(n)}(x) = f(x) + \sqrt{\frac{f_0(x)}{n}} \dot{W}(x), \quad 0 \leq x \leq 1, \quad f \in U_n(f_0), \tag{6}$$

where $W(x)$ is a standard Wiener process. The following statement is proved in [9].

Proposition 1. *For any $\beta > 1/2$ and for f_0 fixed the density experiment with observations $X_1, ..., X_n$ is locally equivalent over $U_n(f_0)$ as $n \to \infty$ to the regression experiment (6) with observation $Y^{(n)}(x)$, $0 \leq x \leq 1$.*

Moreover, this equivalence can be made global. Let $N = n/\log n$, and let $\hat{f}_N = \hat{f}_N(x|X_1, ..., X_N)$ be an estimator of $f(x)$, satisfying

$$(7) \qquad \inf_{f \in \Sigma(\beta, L, b)} P_f^{(n)} \left(\hat{f}_N \in U_n(f) \right) \longrightarrow 1 \quad \text{as } n \to \infty.$$

For $\beta > 1/2$ such an estimator can be defined e.g. by a standard histogram procedure. Let $\bar{f}_N$ be a discretized version with values in $\Sigma(\beta, L, b)$, e.g. a projection of $\hat{f}_N$ onto a finite ϵ-net in $\Sigma(\beta, L, b)$. Consider observations $Y_*^{(n)}(x)$ satisfying the Ito equation:

$$(8) \qquad \dot{Y}_*^{(n)}(x) = f(x) + \left(\frac{\bar{f}_N(x)}{n - N} \right)^{1/2} \dot{W}(x), \quad 0 \le x \le 1.$$

Proposition 2. *For any $\beta > 1/2$ the density experiment with observations $X_1, ..., X_n$ is globally equivalent over $\Sigma(\beta, L, b)$ as $n \to \infty$ to the compound experiment with observations $(X_1, ..., X_N; Y_*^{(n)}(x))$ defined by (8).*

3 Asymptotics of Sup-Norm Risk in Regression

Consider the continuous regression

$$(9) \qquad \dot{Z}^{(n)}(x) = f(x) + \frac{\sigma(x)}{\sqrt{n}} \dot{W}^{(n)}(x), \quad 0 \le x \le 1,$$

where $\sigma = \sigma(x)$ is a given continuous function in x.
First we consider the case when σ is a constant, $\sigma^2 > 0$. In this case the exact asymptotics of the sup–norm risk was found in [5], Donoho [3]. As shown in these papers, if $\hat{f}_n$ is an estimator of f from the observations (9) and if the minimax risk r_n is defined by the expression (1) then for any loss function $w(u)$ we have the following limit result

$$(10) \qquad \lim_{n \to \infty} r_n = w \left(C(\beta, L, \sigma^2) \right)$$

where

$$(11) \qquad C(\beta, L, \sigma^2) = A_\beta \left(\frac{2\sigma^2 L^{1/\beta}}{2\beta + 1} \right)^{\beta/(2\beta+1)}.$$

For the case where $\sigma^2 = \sigma^2(x) > 0$ is a known continuous function the result (10), (11) can be extended as follows.

Theorem 2 . *Let in (9) $\sigma = \sigma(x)$ be a known function continuous in $x \in [0, 1]$, $\sigma(x) < 0$, and let $C(\beta, L, \sigma_*^2)$ be given by (11) where*

$$\sigma_*^2 = \max_{0 \le x \le 1} \sigma^2(x)$$

Then for any $\beta > 0$ and for any loss function $w(u)$ the minimax risk r_n in the regression model (9) satisfies

$$(12) \qquad \lim_{n \to \infty} r_n = w \left(C(\beta, L, \sigma_*^2) \right).$$

The proof of this theorem goes along the lines of [5], [3]. There is only one non–trivial fact which must be taken into account: the sup–norm risk is asymptotically independent of the choice of interval. Let

$$||f||_{[x_1,x_2]} = \sup_{x_1 \le x \le x_2} |f(x)|, \quad 0 \le x_1 < x_2 \le 1.$$

Lemma 1. *Let in (9) σ^2 be a positive constant, let $\hat{f}_n = \hat{f}_n(x|Z^{(n)})$ be an estimator of the regression function f obtained from observations (9), and suppose that loss is measured on an interval $[x_1, x_2]$, $0 \le x_1 < x_2 \le 1$ only. Then the asymptotics of the minimax risk is independent of the interval, i.e. for any loss function $w(u)$ we have*

$$\lim_{n\to\infty} \inf_{\hat{f}_n} \sup_{f \in \Sigma(\beta,L,b)} E_f^{(n)} w\left(\psi_n^{-1} ||\hat{f}_n(\cdot|Z^{(n)}) - f||_{[x_1,x_2]}\right) = w\left(C(\beta, L, \sigma^2)\right). \tag{13}$$

Proof. Rescaling the interval $[x_1, x_2]$ into $[0,1]$ we reduce the n in (9) to $(x_2 - x_1)n$ and turn the Lipschitz constant L into $L(x_2 - x_1)^\beta$ which means that the product

$$C(\beta, L, \sigma^2)\psi_n = A_\beta \left(\frac{2\sigma^2}{2\beta+1} L^{1/\beta} (\log n)/n \right)^{\beta/(2\beta+1)}$$

stays asymptotically unchanged as $n \to \infty$.This proves attainability. For the lower risk bound, note that the reasoning in [5] for $[x_1, x_2] = [0,1]$ shows that additional observations outside $[0,1]$ can be neglected. □

Proof of Theorem 1. Let $\epsilon > 0$ and take an efficient estimator f_n^* in Theorem 2 obtained from observations (9) for $\sigma^2(x) \equiv B_*$. Substitute the observations $Z^{(n)}(x)$ in this estimator by $Y_*^{(n)}(x)$ from (8). Since the preliminary estimator $\bar{f}_N$ takes values in $\Sigma(\beta, L, b)$, we have with probability one $\left\|\bar{f}_N\right\|_\infty \le B_*$. It is clear from the structure of the optimal estimator f_n^* (see [5]) that its risk is monotone in the variance function, i. e. if applied in a model (9) with $\sigma^2(x) \le B_*$ it does not behave worse than for the model it was designed for: $\sigma^2(x) \equiv B_*$. Since the loss function is bounded, we obtain that in the compound experiment (7), (8) the asymptotic minimax risk $w\left(C(\beta, L, B_*)\right)$ is attainable. Proposition 2 then implies the upper bound in (4):

$$\lim_{n\to\infty} \sup r_n \le w\left(C(\beta, L, B_*)\right).$$

To complete the proof note that the lower bound in (13) can be sharpened as follows. If $\beta > 1/2$ then for any $f_0 \in \Sigma(\beta, L, b)$ and for an arbitrary intervall $[x_1, x_2] \subseteq [0,1]$ one has

$$\lim_{n\to\infty} \inf_{\hat{f}_n} \sup_{f \in U_n(f_0)} E_f^{(n)} w(\psi_n^{-1} ||\hat{f}_n - f||_{[x_1,x_2]}) \ge C(\beta, L, \sigma^2). \tag{14}$$

The equality (14) is a direct extension of the corresponding result in [5] and [3] where for σ constant a neighborhood of magnitude $O(\psi_n)$ is used to derive the lower bound. In the case $\beta > 1/2$ this neighborhood is contained in $U_n(f_0)$ for n large (see also (17) in the next section). Now the lower bound $\liminf_{n\to\infty} r_n \ge w\left(C(\beta, L, B_*)\right)$ is a consequence of (14) and Proposition 1. □

4 The case of Hölder exponent $\beta \le 1/2$

In this section we discuss the case $\beta \in (0, 1/2]$. This case is not covered by the previous reasoning: through Theorem 2 is valid for any $\beta > 0$ the argument of equivalence (Propositions 1 and 2) is not applicable. In fact it has been established (see [1]) that asymptotic equivalence does not hold for $\beta < 1/2$. As we show here, the asymptotics of the risk (1) in this case can be obtained from some direct calculations similar to [5]. These will be seen to go through for all $\beta \in (0, 1]$ at once, so we admit some intersection with the previous case here.

Theorem 3. *If $0 < \beta \le 1$ then for any loss function $w(u)$ the minimax risk (1) satisfies*

(15)
$$\lim_{n\to\infty} r_n = w(C_0)$$

where

$$C_0 = \left(B_* L^{1/\beta}(\beta+1)/(2\beta^2)\right)^{\beta/(2\beta+1)}.$$

Remark. In the case $0 < \beta \le 1$ the auxiliary problem (3) has an explicit solution $A_\beta = ((2\beta+1)(\beta+1)/(4\beta^2))^{\beta/(2\beta+1)}$, i. e. the right–hand sides in (4) and (15) coincide.

Proof. The upper bound. Take the kernel $K(u) = (2\beta)^{-1}(\beta+1)(1-|u|^\beta)_+,\ u \in R^1$, and the bandwith $h_n = (C_0\psi_n/L)^{1/\beta}$. For arbitrary small fixed $\epsilon > 0$ define regular grid points $x_k = \epsilon k \psi_n,\ k = 0, \ldots, M$ where $M = M(n) = (\epsilon\psi_n)^{-1}$ is assumed integer. Take $\kappa = ((\beta+1)/(2\beta+1))^{1/\beta}$. To take account of edge effects we put :

$$K_0(X, x_k) = K(h_n^{-1}(X - x_k)) \text{ if } h_n \le x_k < 1 - h_n;$$

$$K_0(X, x_k) = \kappa^{-1} I_{[0,\kappa h_n]}(X - x_k) \text{ if } 0 \le x_k < h_n;$$

$$K_0(X, x_k) = \kappa^{-1} I_{[-\kappa h_n,0]}(X - x_k) \text{ if } 1 - h_n < x_k \le 1.$$

Introduce the kernel estimator

$$f_n^*(x_k) = (nh_n)^{-1} \sum_{i=1}^n K_0(X_i, x_k),\ k = 0, ..., M.$$

Finally, for $x \in (x_{k-1}, x_k)$ define $f_n^*(x)$ as the linear interpolation

$$f_n^*(x) = \alpha(x) f_n^*(x_{k-1}) + (1 - \alpha(x)) f_n^*(x_k), \quad \alpha(x) = (\epsilon\psi_n)^{-1}(x_k - x).$$

Introduce the bias $b_n(x) = E_f^{(n)} f_n^*(x) - f(x),\ 0 \le x \le 1$, and the centered stochastic term $z_n(x) = f_n^*(x) - f(x) - b_n(x)$.

Lemma 2. *Uniformly in $f \in \Sigma(\beta, L, b)$ we have the inequality*

$$\psi_n^{-1} ||b_n||_\infty \le C_0/(2\beta+1) + L\epsilon^\beta.$$

Proof. For $x_k \in [h_n, 1-h_n]$ we have

$$|b_n(x_k)| \leq \int K(u)|f(x_k+h_nu)-f(x_k)|du \leq Lh_n^\beta \int K(u)|u|^\beta du = C_0\psi_n/(2\beta+1).$$

If $x_k \in [0, h_n]$ then

$$|b_n(x_k)| \leq \kappa^{-1}Lh_n^\beta \int_0^\kappa u^\beta du = C_0\psi_n\kappa^\beta/(\beta+1) = C_0\psi_n/(2\beta+1). \ \square$$

Lemma 3. *For arbitrary small $\alpha > 0$ we have*

$$\lim_{n\to\infty} \sup_{f\in\Sigma(\beta,L,b)} P_f^{(n)}\left(\psi_n^{-1}||z_n||_\infty > (1+\alpha)2\beta C_0/(2\beta+1)\right) = 0.$$

Proof. Define the random variables

$$\xi_{ik} = h_n^{-1}K_0(X_i, x_k) - E_f^{(n)}\left[h_n^{-1}K_0(X_i,x_k)\right], \ i = 1,...,n, \ k = 0,...,M.$$

Note that for any $x_k \in [h_n, 1-h_n]$ the variance

$$D_{nk}^2 = \mathrm{Var}_f^{(n)}[\xi_{ik}] = f(x_k)h_n^{-1}\frac{(\beta+1)}{(2\beta+1)}(1+o(1)) \text{ as } n\to\infty,$$

and $E_f^{(n)}[\xi_{ik}^3] = O(h_n^{-2})$ as $n\to\infty$ uniformly in $f \in \Sigma(\beta, L, b)$. Note that for n large

$$\left((\psi_n\sqrt{n}/D_{nk})(1+\alpha)2\beta C_0/(2\beta+1)\right)^2 =$$

$$(B_*/f(x_k))(1+\alpha)^2\frac{2}{(2\beta+1)}\log n \ (1+o(1)) > (2+\alpha)\log M.$$

Put $A_n = ((2+\alpha)(\log M))^{1/2}$. The Chebyshev exponential inequality yields

$$\lim_{n\to\infty} P_f^{(n)}\left(\psi_n^{-1}\max_{x_k\in[h_n,1-h_n]} z_n(x_k) > (1+\alpha)2\beta C_0/(2\beta+1)\right)$$

$$\leq \sum_{x_k\in[h_n,1-h_n]} P_f^{(n)}\left((nD_{nk}^2)^{-1/2}\sum_{i=1}^n \xi_{ik} > A_n\right)$$
$$\leq \sum_{x_k\in[h_n,1-h_n]} \exp(-cA_n)E_f^{(n)}\left[\exp(c\xi_{1k}(nD_{nk}^2)^{-1/2}\right]^n$$

$$\leq M\exp(-cA_n)\left(1+\frac{c^2}{2n}(1+o(1)\right)^n \text{ as } n\to\infty.$$

Now take $c = A_n$ and let $\delta > 0$ be arbitrarily small. The above latter expression does not exceed $M\cdot\exp(-A_n^2+\frac{1}{2}A_n^2(1+\delta)) = M\cdot M^{-(1+\alpha/2)(1-\delta)} = M^{-\alpha/2}M^{(1+\alpha/2)\delta}$ which is vanishing as $n\to\infty$ if δ is small enough. For the same reasons the probability

$P_f^{(n)}\left(\psi_n^{-1}\min_{x_k\in[h_n,1-h_n]} z_n\left(x_k\right) < -(1+\alpha)2\beta C_0/(2\beta+1)\right)$ is vanishing as well. Note that there are only a bounded number of points $x_k \in [0,h_n)\cup(1-h_n,1]$ and asymptotically as $n\to\infty$ they do not influence the value of $||z_n||_\infty$. Thus the lemma follows. $\square$

The lower bound. It suffices to prove that for an arbitrary estimator $\hat{f}_n$ and for any small $\alpha>0$

(16) $$\liminf_{n\to\infty} \sup_{f\in\Sigma(\beta,L,b)} P_f^{(n)}\left(||\hat{f}_n - f||_\infty \ge (1-\alpha)C_0\psi_n\right) = 1.$$

Take a small value $\epsilon=\epsilon(\alpha)>0$, the final choice of ϵ will be made below. Let $f_*\in\Sigma(\beta,L,b)$ be such that $f_*(x)=f_0$ is a constant in $x\in[t_1,t_2], t_2-t_1=\epsilon$, and $f_0 = B_*/(1+\epsilon)$. Introduce a family of functions

(17) $$f(x;\theta) = f(x;\theta_1,...,\theta_M) = f_*(x) + Lh_n^\beta \sum_{j=1}^M \theta_j g(h_n^{-1}(x-h_n a_j)), \quad 0\le x\le 1,$$

where $a_1 = x_1,\ \ a_{j+1}-a_j = 2(1+1/\epsilon),\quad j=1,...,M,\quad M=M(n)=[n^{1/((2\beta+1)\,(1+\epsilon))}]$, $\theta=(\theta_1,...,\theta_M)\in K=[-1,1]^M$; the "basic" function is

$$g(u) = \left(1-|u-1|^\beta\right)_+ - \epsilon\left(1-|\epsilon u-(2\epsilon+1)|^\beta\right)_+,\ u\in R^1.$$

As is easily seen, $\int g=0,\ \int g^2 = (1+\epsilon)\,4\beta^2/(\beta+1)(2\beta+1)$, and for ϵ small enough and n sufficiently large $f(x;\theta)\in\Sigma(\beta,L,b)$ for $\theta\in K$. The density $f(x;\theta)$ differs from $f_*(x)$ only in the interval $[t_1,t_1+2(1+1/\epsilon)Mh_n]\subseteq[t_1,t_2]$ for n large since $Mh_n\to 0$ as $n\to\infty$ for any ϵ fixed. Put for shortness $P_{f(\cdot;\theta)}^{(n)} = P_\theta^{(n)}$ and $E_{f(\cdot;\theta)}^{(n)} = E_\theta^{(n)}$. The log–likelihood ratio is given by

$$\Lambda_n(\theta) = \log\left(dP_\theta^{(n)}/P_0^{(n)}\right) = \sum_{i=1}^n \log\left(1 + f_0^{-1}Lh_n^\beta\sum_{j=1}^M \theta_j g(h_n^{-1}(X_i-h_n a_j))\right).$$

Define $\Delta_j = f_0^{-1}Lh_n^\beta\sum_{i=1}^n g(h_n^{-1}(X_i-h_n a_j)),\quad j=1,...,M$. Note that

$$E_0^{(n)}\Delta_j = 0;\ E_0^{(n)}\Delta_j^2 = (1+\epsilon)^2\,(2/(2\beta+1))\log n = 2(1+\epsilon)^3\log M,$$

$$E_0^{(n)}\Delta_j\Delta_k = 0 \text{ if } \ j\ne k,\ j,k=1,...,M.$$

The Taylor expansion implies that $\Lambda_n(\theta)$ can be approximated by

$$\widetilde{\Lambda}_n(\theta) = \sum_{j=1}^M\left(\theta_j\Delta_j - \frac{1}{2}\theta_j^2 E_0^{(n)}\Delta_j^2\right) = \sum_{j=1}^M\left(\theta_j\Delta_j - \theta_j^2(1+\epsilon)^3\log M\right).$$

Lemma 4. $\lim_{n\to\infty} \sup_{\theta\in K} E_0^{(n)} \left|\exp(\Lambda_n(\theta)) - \exp(\widetilde{\Lambda}_n(\theta))\right| = 0$.

We skip the proof of this lemma, just noting that

$$E_0^{(n)} \left|\exp(\Lambda_n(\theta)) - \exp(\widetilde{\Lambda}_n(\theta))\right| = E_\theta^{(n)} \left|1 - \exp(\widetilde{\Lambda}_n(\theta) - \Lambda_n(\theta))\right|,$$

and $\widetilde{\Lambda}_n(\theta) - \Lambda_n(\theta) \to 0$ as $n \to \infty$ in $P_\theta^{(n)}$-probability uniformly in $\theta \in K$. Moreover these random variables do have exponentially fast decreasing distribution tails. Calculations are based on the Chebyshev exponential inequality.

Though the random variables Δ_j are non–correlated under $P_0^{(n)}$ they are dependent via the sample $X_1, ..., X_n$. But they are conditionally independent given the number of sample points in the support of each function $g(h_n^{-1}(\cdot - h_n a_j))$. Define

$$\nu_j = \#\left\{i,\, 1 \le i \le n:\, X_i \in [h_n a_j, h_n a_{j+1}]\right\}, \quad j = 1, ..., M.$$

Introduce the events

$$\mathcal{A}_{jn} = \left\{|\nu_j - n\pi_0| \ge n^{\epsilon/4}(n\pi_0)^{1/2}\right\}, \quad j = 1, ..., M, \text{ where } \pi_0 = 2(1 + 1/\epsilon) f_0 h_n.$$

Note that the joint $P_\theta^{(n)}$-distribution of $(\nu_1, ..., \nu_M)$ is independent of $\theta \in K$, and $P_\theta^{(n)}\left(\bigcup_{j=1}^M \mathcal{A}_{jn}\right) \to 0$ as $n \to \infty$, or

$$\lim_{n\to\infty} \inf_{\theta\in K} P_\theta^{(n)}(\mathcal{B}_n) = 1 \text{ where } \mathcal{B}_n = \bigcap_{j=1}^M \bar{\mathcal{A}}_{jn}. \tag{18}$$

The crucial point of the proof is that Δ_j are $P_0^{(n)}$-conditionally independent given $\nu_1, ..., \nu_M$, i. e.

$$P_0^{(n)}(\Delta_1 \le t_1, ..., \Delta_M \le t_M \mid \nu_1 = n_1, ..., \nu_M = n_M) = \tag{19}$$

$$= \prod_{j=1}^M P_0^{(n)}(\Delta_j \le t_j \mid \nu_j = n_j).$$

Direct calculations show that if $n_j : |n_j - n\pi_0| \le n^{\epsilon/4}(n\pi_0)^{1/2}$ then

$$E_0^{(n)}\left(\exp(\theta_j\Delta_j - \theta_j^2(1+\epsilon)^3(\log M)) \mid \nu_j = n_j\right) = 1 + o(M^{-1}) \tag{20}$$

where $M\ o(M^{-1}) \to 0$ as $n \to \infty$ uniformly in $|\theta_j| \le 1$.
Now we are ready to prove (16). We omit those details which are similar to the gaussian case in [5]. First, standard arguments show that (16) is equivalent to

$$\liminf_{n\to\infty} \sup_{\theta\in K} P_\theta^{(n)}\left(||\hat{\theta}_n - \theta||_M \ge 1 - \alpha\right) = 1 \tag{21}$$

where $\hat{\theta}_n = (\hat{\theta}_{n1}, ..., \hat{\theta}_{nM})$ is an arbitrary estimator of $\theta = (\theta_1, ..., \theta_M)$, $||\theta||_M = \max_{1\le j\le M} |\theta_j|$. Put $\mathcal{C}_n = \left\{||\hat{\theta}_n - \theta||_M < 1-\alpha\right\}$. Applying Lemma 4 and (18), we obtain that

$$\sup_{\theta\in K} P_\theta^{(n)}\left(\bar{\mathcal{C}}_n\right) \ge 1-\epsilon-2^{-M}\int_K E_0^{(n)}\left(\exp(\tilde{\Lambda}_n(\theta))I(\mathcal{B}_n)I(\mathcal{C}_n)\right)d\theta. \tag{22}$$

As in the gaussian case (see [5]), the minimal value of the right hand side of (22) is attained if we take $\hat{\theta}_n = \tau_n^* = (\tau_{n1}^*, ..., \tau_{nM}^*)$; $\tau_{nj}^* = \widetilde{\Delta}_j$ if $|\widetilde{\Delta}_j| \le \alpha$, and $\tau_{nj}^* = \alpha$ if $\widetilde{\Delta}_j > \alpha$, $\tau_{nj}^* = -\alpha$ if $\widetilde{\Delta}_j < -\alpha$ where $\widetilde{\Delta}_j = \Delta_j/(2(1+\epsilon)^3\log M)$. Define a set of integers

$$\mathcal{N} = \left\{(n_1, ..., n_M):\ |n_j - n\pi_0| \le n^{\epsilon/4}(n\pi_0)^{1/2},\ j=1,...,M\right\}.$$

Let $\eta_j = \frac{1}{2}\exp\left\{\theta_j\Delta_j - \theta_j^2(1+\epsilon)^3\log M\right\}$, $\kappa_j = I(|\tau_{nj}^* - \theta_j| < 1-\alpha)$. For n large (22) implies that

$$\begin{aligned}
\sup_{\theta\in K} P_\theta^{(n)}\left(\bar{\mathcal{C}}_n\right) &\ge \\
&\ge 1-\epsilon-E_0^{(n)}I(\mathcal{B}_n)\left[\int_K\prod_{j=1}^M \eta_j\ \kappa_j\ d\theta\right] \\
&= 1-\epsilon
\end{aligned}$$

$$-\sum_{(n_1,...,n_M)\in\mathcal{N}} P_0^{(n)}(\nu_1 = n_1, ..., \nu_M = n_M)\ E_0^{(n)}\left[\int_K\prod_{j=1}^M \eta_j\ \kappa_j\ d\theta \mid \nu_1 = n_1, ..., \nu_M = n_M\right].$$

In view of (19) and (20) each conditional expectation is a product

$$\prod_{j=1}^M E_0^{(n)}\left[\int_{-1}^1 \eta_j\ \kappa_j d\theta_j \mid \nu_1 = n_1, ..., \nu_M = n_M\right] \le$$

$$\le \prod_{j=1}^M\left(1+o(M^{-1}) - \int_{-1}^1 E_0^{(n)}\left[\eta_j(1-\kappa_j) \mid \nu_1 = n_1, ..., \nu_M = n_M\right]\ d\theta_j\right) \tag{23}$$

Here on the event $\{\tilde{\Delta}_j \le \alpha/4\}$ we have

$$\eta_j(1-\kappa_j) =$$

$$\begin{aligned}
&= \int_{-1}^1 \frac{1}{2}(1-\kappa_j)\exp\left(-\frac{1}{2}(E_0^{(n)}\Delta_j^2)(\theta_j-\tilde{\Delta}_j)^2\right)\exp\left(\frac{1}{2}(E_0^{(n)}\Delta_j^2)\tilde{\Delta}_j^2\right)d\theta_j \\
&\ge \frac{1}{2}\int_{-1-\tilde{\Delta}_j}^{1-\tilde{\Delta}_j} I(|y|\ge 1-\alpha)\exp\left(-\frac{1}{2}(E_0^{(n)}\Delta_j^2)y^2\right)dy \\
&\ge \frac{1}{2}\int_{1-\alpha}^{1-\alpha/2}\exp\left(-\frac{1}{2}(E_0^{(n)}\Delta_j^2)y^2\right)dy \ge \frac{\alpha}{4}\exp\left(-(1+\epsilon)^3(\log M)(1-\alpha/2)^2\right).
\end{aligned}$$

Hence

$$\int_{-1}^{1} E_0^{(n)} \left[\eta_j(1-\kappa_j) \mid \nu_1 = n_1, ..., \nu_M = n_M\right] \; d\theta_j$$
$$\geq \quad (\alpha/4)M^{-\alpha_1}, \quad \alpha_1 = (1+\epsilon)^3(1-\alpha/2)^2.$$

and (23) does not exceed

$$\left(1 + o(M^{-1}) - (\alpha/4)M^{-\alpha_1}\right)^M .$$

Choose ϵ such that $\alpha_1 = (1-\alpha/2)$. Under this choice for all M large one has the inequality $(1+o(M^{-1}) - (\alpha/4)M^{-\alpha_1})^M \leq \epsilon$. Hence

$$\sup_{\theta \in K} P_\theta^{(n)}\left(\bar{\mathcal{C}}_n\right) \geq 1 - \epsilon - \epsilon \sum_{(n_1,...,n_M)\in\mathcal{N}} P_0^{(n)}\left(\nu_1 = n_1, ..., \nu_M = n_M\right) \geq 1 - 2\epsilon. \;\square$$

References

[1] Brown, L. D. and Zhang, C.-H. (1995). Nonparametric density and regression are not asymptotically equivalent to the nonparametric white noise model when the smoothness index is $< \frac{1}{2}$. Manuscript.

[2] Devroye, L. , Gyorfi, L. (1985) *Nonparametric Density Estimation: the L_1 -view.* John Wiley, N. Y.

[3] Donoho, D. (1992) Asymptotic minimax risk (for sup–norm loss): solution via optimal recovery. Probab. Theor. Rel. Fields **99** 145-170.

[4] Ibragimov, I.A. , Khasminskii, R.Z. (1981) *Statistical Estimation: Asymptotic Theory.* Springer, New York.

[5] Korostelev, A.P. (1993) Exact asymptotically minimax estimator for nonparametric regression in uniform norm. Theory Probab. Appl. **38**, no. 4, 775–782

[6] Le Cam, L. (1985). Sur l'approximation de familles de mesures par des familles gaussiennes. *Ann. Inst. Henri Poincaré* **21** (3) 225-287

[7] Le Cam , L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

[8] Low, M. (1992). Renormalization and white noise approximation for nonparametric functional estimation problems. *Ann. Statist.* **20** 545- 554

[9] Nussbaum, M. (1995). Asymptotic Equivalence of Density Estimation and White Noise. To appear, Ann. Statist.

[10] Pinsker, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise. *Problems Inform. Transmission* (1980) 120-133

Dept. of Mathematics
Wayne State University
Detroit, MI 48202
USA
e-mail apk@math.wayne.edu

Weierstrass Institute
Mohrenstr. 39
D-10117 Berlin
Germany
e-mail nussbaum@iaas-berlin.d400.de